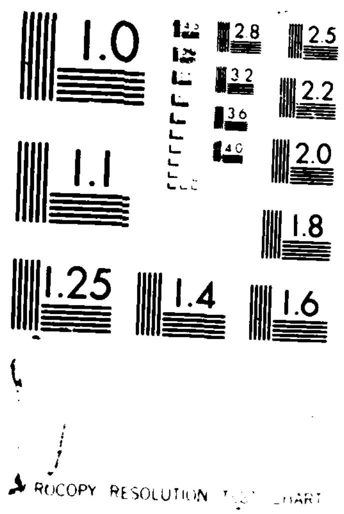
AD-A186 025 STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF REGRESSION VARIABLES WHEN T (U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Y WU JUN 87 TR-87-15 1/1

UNCLASSIFIED AFOSR-TR-87-1245 F49620-85-C-0008 F/G 12/3 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
1.1
2.0
1.8
1.25
1.4
1.6
ROCOPY RESOLUTION
HART

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

AD-A186 025

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 87-1245	2. GOVT ACCESSION NO	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Strong consistency of estimation of number of regression variables when the errors are independent and their expectations are not equal to each other.		5. TYPE OF REPORT & PERIOD COVERED Final ~~technical~~ - June 1987
		6. PERFORMING ORG. REPORT NUMBER 87-15
7. AUTHOR(s) Yuehua Wu		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis University of Pittsburgh, 515 Thackeray Hall Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332 NM		12. REPORT DATE June 1987
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same as 11		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 16 1987 S D

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear model, model selection, regression coefficient, strong consistency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider the linear regression model $y_i = x_i'\beta + e_i$, $i = 1,2,\ldots$, where $\{x_i\}$ is a sequence of known p-vectors, $\beta' = (\beta_1,\ldots,\beta_p)$ is an unknown p-vector, known as regression coefficients, $\{e_i\}$ is a sequence of random errors. It is of interest to test the hypothesis H_k: $\beta_{k+1} = \ldots = \beta_p = 0$, $k = 0,1,\ldots,p$. We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support $[a_1,a_2]$. Under certain

DD FORM 1 JAN 73 1473

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

conditions, we obtain the strongly consistent estimate of the number of k for which $\beta_k \neq 0$ and $\beta_{k+1} = \ldots \beta_p = 0$, by using the information theoretical criteria.

AFOSR-TR- 87-1245

STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF REGRESSION VARIABLES WHEN THE ERRORS ARE INDEPENDENT AND THEIR EXPECTATIONS ARE NOT EQUAL TO EACH OTHER*

Yuehua Wu
Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis

University of Pittsburgh

87 9 24 04

AFOSR-TR- 87-1245

STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF REGRESSION VARIABLES WHEN THE ERRORS ARE INDEPENDENT AND THEIR EXPECTATIONS ARE NOT EQUAL TO EACH OTHER*

Yuehua Wu
Center for Multivariate Analysis
University of Pittsburgh

June 1987

Technical Report No. 87-15

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*Research partially supported by the Air Force Office of Scientific Research (AFSC) under contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation heron.

Abstract

Consider the linear regression model $y_i = x_i'\beta + e_i$, $i = 1, 2, \ldots,$ where $\{x_i\}$ is a sequence of known p-vectors, $\beta' = (\beta_1, \ldots, \beta_p)$ is an unknown p-vector, known as regression coefficients, $\{e_i\}$ is a sequence of random errors. It is of interest to test the hypothesis H_k: $\beta_{k+1} = \ldots = \beta_p = 0$, $k = 0, 1, \ldots, p$. We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support $[a_1, a_2]$. Under certain conditions, we obtain the strongly consistent estimate of the number k for which $\beta_k \neq 0$ and $\beta_{k+1} = \ldots = \beta_p = 0$, by using the information theoretical criteria.

1. Introduction

Consider the linear model

$$y_i = x_i'\beta + e_i, \; i=1,2, \; \ldots \; ,n, \tag{1}$$

where x_i's are experiment points, $\beta = (\beta_1, \; \ldots \; ,\beta_n)'$ is the regression coefficient vector to be estimated, and e_i's are random errors. In the usual linear regression model it is assumed that the random errors have vanishing expectations and common variance. In this case, the famous least square estimation (LSE) method plays an important role in making statistical inference upon the regression coefficient vector β. In the literature, there are a lot of papers concerning with the LSE and many important results are obtained (a part of work refers to [1],[2] and [3]). However the unbiasedness and consistency (even the weak one) of LSE strongly depend on the assumption that the expectations of errors are zero, and this assumption is not realistic sometimes. It is of interest to find a consistent estimates of the regression coefficients when the expectations of errors are not equal to each other. In [4] two methods for finding consistent estimates of the regression coefficient vector β are proposed.

The first method is to use the measure

$$Q_n(\beta) = \max_{1\le i\le n} (y_i - x_i'\beta) - \min_{1\le i\le n} (y_i - x_i'\beta)$$

The estimator $\hat{\beta}_n$ of β is defined as the vector which minimizes $Q_n(\beta)$. The estimate $\hat{\beta}_n$ is temporarily called MD estimate of β in [4] (the estimate based on the Maximum Difference between residuals).

The second method is to use the measure

$$\bar{Q}_n(\beta) = \max_{1\le i\le n} |y_i - x_i'\beta|$$

Denote by $\bar{\beta}_n$ the value of β which minimizes $\bar{Q}_n(\beta)$. Also, $\bar{\beta}_n$ is temporarily called MA eatimate of β (the estimate based on the Maximum Absolute values of residuals)

Under certain conditions, both $\hat{\beta}_n$ and $\bar{\beta}_n$ are shown to be strongly consistent in [4].

Now let us consider the hypotheses

$$H_k: \beta_{k+1} = \beta_{k+2} = \ldots = \beta_p = 0 \text{ and } \beta_k \neq 0$$
$$k = 0, 1, \ldots, p-1.$$

It is of interest to determine the true model H_k by using the model selection criteria. Denote by $\hat{\beta}_{kn} = (\hat{\beta}_{k1n}, \ldots, \hat{\beta}_{kkn}, 0, \ldots, 0)'$ the vector which minimizes $Q_n(\beta)$ under the restriction $\beta_{k+1} = \ldots = \beta_p = 0$ and denoted by $\bar{\beta}_{kn} = (\bar{\beta}_{k1n}, \ldots, \bar{\beta}_{kkn}, 0, \ldots, 0)'$ the vector which minimizes $\bar{Q}_n(\beta)$ under the restriction $\beta_{k+1} = \ldots = \beta_p = 0$. Write

$$\hat{Q}_k = Q_n(\hat{\beta}_{kn})$$

and

$$\bar{Q}_k = \bar{Q}_n(\bar{\beta}_{kn})$$

Choose a sequence of constants C_n, satisfying certain conditions which will be specified later, and define

$$\hat{R}_k = \hat{Q}_k + kC_n$$

and

$$\bar{R}_k = \bar{Q}_k + kC_n$$

Choose

$$\hat{k} = \text{ArgMin}\{\hat{R}_k : k \in \{0, \ldots, p\}\}$$

and

$$\bar{k} = \text{ArgMin}\{\bar{R}_k : k \in \{0, \ldots, p\}\}$$

where ArgMin denote the index which minimizes the quantities following the symbol ArgMin.

In this paper we shall consider the consistency of $\hat{k}$ and $\bar{k}$ to the true model k_0.

2. Consistency of $\hat{k}$

In this section, we make the following general assumptions:

Assumption 1. The errors e_i, $i = 1,2, \ldots$ are independent.

Assumption 2. $P\{e_n \bar{\in} [a_1, a_2]\} = 0$ and there is a positive constant Δ such that for any $\varepsilon > 0$ and any n, we have

$$P\{e_n \in [a_1, a_1 + \varepsilon]\} \geq \Delta\varepsilon$$

and

$$P\{e_n \in [a_2 - \varepsilon, a_2]\} \geq \Delta\varepsilon.$$

Assumption 3. For any $a > 0$, there exists a positive constant C such that for any vector $\alpha \neq \mathbf{0}$ it follows that

$$\#\{i \leq n, \ |\ell(x_i) - \ell(\alpha)| < a\} \geq Cn$$

for large n, hereafter $\ell(\alpha) = \alpha / |\alpha|$

Assumption 4. There exists a positive constant m such that

$$|x_i| > m, \qquad \text{for } i = 1,2, \ldots$$

Now let us estimate $Q_n(\hat{\beta}_n)$. Define

$$E_n^{(1)} = \{i \le n,\ -x_i'(\hat{\beta}_n - \beta) > 0\}$$

$$E_n^{(2)} = \{i \le n,\ x_i'(\hat{\beta}_n - \beta) > 0\}$$

Split $S_p = \{ x \in R^p: |x| = 1\}$ into d disjoint parts $\Sigma_1, \ldots, \Sigma_d$ such that $\forall$ $x, y \in \Sigma_j$, $x'y > 3/4$. Let $\gamma_j \in \Sigma_j$, $j = 1, \ldots, d$. Define $E_n^j = \{i \le n, \ell(x_i)'\gamma_j > 3/4\}$, $j = 1, \ldots, d$. By Assumtion 3, there exists $\delta_1 > 0$ such that

$$\#(E_n^j) \ge \delta_1 n, \qquad j = 1,2, \ldots, d.$$

It is easy to see that $-\ell(\hat{\beta}_n - \beta) \in \Sigma_j$ and $i \in E_n^j$ implies that

$$-x_i'\ell(\hat{\beta}_n - \beta) > 0, \qquad \text{i.e. } i \in E_n^{(1)}$$

and that $\ell(\hat{\beta}_n - \beta) \in \Sigma_j$ and $i \in E_n^j$ implies that

$$x_i'\ell(\hat{\beta}_n - \beta) > 0, \qquad \text{i.e. } i \in E_n^{(2)}$$

Take r_n satisfying

$$r_n \to 0 \quad \text{and} \quad nr_n/\log n \to \infty,$$

we have

$$P(Q_n(\hat{\beta}_n) \le a_2 - a_1 - 2r_n)$$

$$\le P(\max_{i\varepsilon E_n^{(1)}} e_i \le a_2 - r_n) + P(\min_{i\varepsilon E_n^{(2)}} e_i \ge a_1 + r_n)$$

$$\le \sum_{j=1}^{d} P(\max_{i\varepsilon E_n^{(1)}} e_i \le a_2 - r_n, -\ell(\hat{\beta}_n - \beta) \ \varepsilon \ \Sigma_j)$$

$$+ \sum_{j=1}^{d} P(\min_{i\varepsilon E_n^{(2)}} e_i \ge a_1 + r_n, \ell(\hat{\beta}_n - \beta) \ \varepsilon \ \Sigma_j)$$

$$\le \sum_{j=1}^{d} P(\max_{i\varepsilon E_n^{j}} e_i \le a_2 - r_n, -\ell(\hat{\beta}_n - \beta) \ \varepsilon \ \Sigma_j)$$

$$+ \sum_{j=1}^{d} P(\min_{i\varepsilon E_n^{j}} e_i \ge a_1 + r_n, \ell(\hat{\beta}_n - \beta) \ \varepsilon \ \Sigma_j)$$

$$\le \sum_{j=1}^{d} P(\max_{i\varepsilon E_n^{j}} e_i \le a_2 - r_n)$$

$$+ \sum_{j=1}^{d} P(\min_{i\varepsilon E_n^{j}} e_i \ge a_1 + r_n)$$

$$\le 2d(1-\Delta r_n)^{\delta_1 n} \le 2de^{-\Delta r_n \delta_1 n} \le 2d/n^2$$

for large n. By Borel-Cantelli Lemma we have

$$Q_n(\hat{\beta}_n) \ge a_2 - a_1 - 2r_n, \qquad \text{a.s.}$$

when n is large enough.

Let k_0 be the index of the true model and let β_0 be the true parameter. Then obviously we have ,for $p \ge k \ge k_0$

$$Q_n(\hat{\beta}_n) = Q_n(\hat{\beta}_{pn}) \le Q_n(\hat{\beta}_{kn})$$

$$\le Q_n(\hat{\beta}_{k_0 n}) \le Q_n(\beta_0) \le a_2 - a_1$$

Thus

$$0 \le Q_n(\hat{\beta}_{k_0 n}) - Q_n(\hat{\beta}_{kn}) \le 2r_n, \qquad p \ge k \ge k_0$$

If we take C_n such that $C_n \to 0$, $C_n/r_n \to \infty$, then for $k > k_0$

$$\hat{R}_k - \hat{R}_{k_0} = (k - k_0)C_n + Q_n(\hat{\beta}_{kn}) - Q_n(\hat{\beta}_{k_0 n}) > 0, \tag{2}$$

for all large n.

Next, we consider the case of $k < k_0$. Denote

$$\eta = |\beta_{k_0}| > 0$$

and define

$$E_n^+ = \{i \le n, |\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0)| < 1/2\}$$

$$E_n^- = \{i \le n, |\ell(x_i) - \ell(\hat{\beta}_{kn} - \beta_0)| < 1/2\}$$

Split S_p into b disjoint parts $\Pi_1, \ldots, \Pi_b$ such that $\forall$ x, y ε Π_i, $|x - y| <$ 1/4. Let $\xi_j \; \varepsilon \; \Pi_j$, j= 1, . . . , b. Define

$$F_n^j = \{i \leq n, |\ell(x_i) - \xi_j| < 1/4\}, \qquad j = 1, \ldots, b.$$

By Assumption 3, there exists $\delta_2 > 0$ such that

$$\#(F_n^j) \geq \delta_2 n, \qquad j = 1, 2, \ldots, b.$$

It is easy to see that

$$-\ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j \quad \text{and} \quad i \in F_n^j$$

which implies that

$$|\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0)| < 1/2, \qquad \text{i.e. } i \in E_n^+.$$

Also,

$$\ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j \quad \text{and} \quad i \in F_n^j,$$

which implies that

$$|\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0)| < 1/2, \quad \text{i.e. } i \in E_n^-$$

For $i \in E_n^+$, we have

$$x_i'(\hat{\beta}_{kn} - \beta_0) = |x_i||\hat{\beta}_{kn} - \beta_0|\ell(x_i)'\ell(\hat{\beta}_{kn} - \beta_0)$$

$$\leq -m\eta(1 - 1/2) = -m\eta/2.$$

Similarly for $i \in E_n^-$, we have

$$x_i'(\hat{\beta}_{kn} - \beta_0) \geq m\eta/2.$$

Hence

$$Q_n(\hat{\beta}_{kn}) \geq \max_{i \in E_n^+} e_i - \min_{i \in E_n^-} e_i + m\eta$$

Thus

$$P(Q_n(\hat{\beta}_{kn}) \leq a_2 - a_1 + m\eta/2)$$

$$\leq P(\max_{i \in E_n^+} e_i \leq a_2 - m\eta/4) + P(\min_{i \in E_n^-} e_i \geq a_1 + m\eta/4)$$

$$\leq \sum_{j=1}^{b} P(\max_{i \in E_n^+} e_i \leq a_2 - m\eta/4,\ -\ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j)$$

$$+ \sum_{j=1}^{b} P(\min_{i \in E_n^-} e_i \geq a_1 + m\eta/4,\ \ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j)$$

$$\leq \sum_{j=1}^{b} P(\max_{i \in F_n^j} e_i \leq a_2 - m\eta/4,\ -\ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j)$$

$$+ \sum_{j=1}^{b} P(\min_{i \varepsilon F_n^j} e_i \geq a_1 + m\eta/4,\ \ell(\hat{\beta}_{kn} - \beta_0) \varepsilon \Pi_j)$$

$$\leq \sum_{j=1}^{b} P(\max_{i \varepsilon F_n^j} e_i \leq a_2 - m\eta/4)$$

$$+ \leq \sum_{j=1}^{b} P(\min_{i \varepsilon F_n^j} e_i \geq a_1 + m\eta/4)$$

$$\leq 2b(1 - \Delta m\eta/4)^{\delta_2 n} \leq 2be^{-\Delta m\eta \delta_2 n/4} \leq 2b/n^2$$

for large n. By Borel–Cantelli Lemma, we have, with probability one,

$$Q_n(\hat{\beta}_n) \geq a_2 - a_1 + m\eta/2, \qquad \text{for all large } n.$$

Thus for $k < k_0$, we have

$$\hat{R}_k - \hat{R}_{k_0} = Q_n(\hat{\beta}_{kn}) - Q_n(\hat{\beta}_{k_0 n}) - (k_0 - k)C_n$$

$$\geq m\eta/2 - (k_0 - k)C_n > 0, \tag{3}$$

for large n, since $C_n \to 0$.

(2) and (3) imply that $\hat{k}$ is strongly consistent. Summarize the above arguments, we get the following theorem.

Theorem 1. Choose C_n satisfying

(i) $C_n \to 0$,

(ii) $nC_n/\log n \to \infty$.

Suppose the four Assumptions given at the beginning of this section are true, then $\hat{k} \to k$, a.s.

Proof. Use the arguments given before. We only need to note that for any sequence of C_n satisfying (i) and (ii), we can always choose r_n such that

(i)' $r_n/C_n \to 0$,

(ii)' $nr_n/\log n \to \infty$.

Q. E. D.

3. Consistency of $\bar{k}$

In this section, we shall make the following general assumptions:

Assumptiom 1'. The error e_i, $i = 1, 2, \ldots$, are independent,

Assumptiom 2. $|a_1| < a_2$, $\forall\, n \Rightarrow P(e_n \tilde{\varepsilon} [a_1, a_2]) = 0$ there is a positive constant $\tilde{\Delta}$ such that for any $\varepsilon > 0$ and for any n, we have

$$P(e_n \varepsilon [a_2 - \varepsilon, a_2]) \geq \tilde{\Delta}\varepsilon.$$

Assumptiom 3'. Same as Assumptiom 3 in Section 2.

Assumptiom 4'. There exists a positive constant $\tilde{m}$ such that

$$|x_i| > \tilde{m}, \quad \text{for } i = 1, 2, \ldots$$

Now let us estimate $\bar{Q}_n(\bar{\beta}_n)$. Define

$$E_n = \{i \le n,\ x_i'(\beta - \bar{\beta}_n) > 0\}$$

Split S_p into $\tilde{d}$ disjoint parts $\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_{\tilde{d}}$ such that $\forall$ x, y ε $\tilde{\Sigma}_j$, $x'y > 3/4$. Let $\tilde{\gamma}_j \ \varepsilon \ \tilde{\Sigma}_j$, $j = 1, \ldots, \tilde{d}$. Define $\bar{E}_n^j = \{i \le n,\ \ell(x_i)'\tilde{\gamma}_j > 3/4\}$, $j = 1, \ldots, \tilde{d}$. By Assumption 3', there exists $\tilde{\delta}_1 > 0$ such that

$$\#(\bar{E}_n^j) \ge \tilde{\delta}_1 n, \qquad j = 1, \ldots, \tilde{d}$$

It is easy to see that

$$-\ell(\bar{\beta}_n - \beta) \ \varepsilon \ \tilde{\Sigma}_j \quad \text{and} \quad i \ \varepsilon \ \bar{E}_n^j$$

imply that $x_i'\ell(\beta - \bar{\beta}_n) > 0$, i. e. i ε E_n.

Take r_n satisfying

$$r_n \to 0, \qquad nr_n/\log n \to \infty$$

We have

$$P(\bar{Q}_n(\bar{\beta}_n) \le a_2 - r_n) \le P(\max_{i \varepsilon E_n} e_i \le a_2 - r_n)$$

$$\le \sum_{j=1}^{\tilde{d}} P(\max_{i \varepsilon E_n} e_i \le a_2 - r_n,\ \ell(\beta - \bar{\beta}_n) \ \varepsilon \ \tilde{\Sigma}_j)$$

$$\le \sum_{j=1}^{\tilde{d}} P(\max_{i \varepsilon \bar{E}_n^j} e_i \le a_2 - r_n,\ \ell(\beta - \bar{\beta}_n) \ \varepsilon \ \tilde{\Sigma}_j)$$

$$\leq \sum_{j=1}^{\tilde{d}} P(\max_{i \in \bar{E}_n^j} e_i \leq a_2 - r_n)$$

$$\leq \tilde{d}(1 - \tilde{\Delta} r_n)^{\tilde{\delta}_1 n} \leq \tilde{d} e^{-\tilde{\Delta} r_n \tilde{\delta}_1 n} \leq \tilde{d}/n^2$$

for large n. By Borel–Cantelli Lemma we have

$$\bar{Q}_n(\bar{\beta}_n) \geq a_2 - r_n, \qquad \text{a.s.}$$

when n is large enough.

Let k_0 be the index of the true model and let β_0 be the true parameter. Then obviously we have for $p \geq k \geq k_0$

$$\bar{Q}_n(\bar{\beta}_n) = \bar{Q}_n(\bar{\beta}_{pn}) \leq \bar{Q}_n(\bar{\beta}_{kn})$$

$$\leq \bar{Q}_n(\bar{\beta}_{k_0 n}) \leq \bar{Q}_n(\beta_0) \leq a_2$$

Thus

$$0 \leq \bar{Q}_n(\bar{\beta}_{k_0 n}) - \bar{Q}_n(\bar{\beta}_{kn}) \leq r_n, \qquad p \geq k \geq k_0.$$

If we take C_n such that

$$C_n \to 0, \qquad C_n/r_n \to \infty$$

then for $k > k_0$

$$\bar{R}_k - \bar{R}_{k_0} = (k - k_0)C_n + \bar{Q}_n(\bar{\beta}_{kn}) - \bar{Q}_n(\bar{\beta}_{k_0 n}) > 0 \tag{4}$$

for all large n.

Next, we consider the case of $k < k_0$. Denote

$$\tilde{\eta} = |\beta_{k_0}| > 0$$

and define

$$\bar{E}_n = \{i \le n,\ \ell(x_i)'\ell(\bar{\beta}_{kn} - \beta_0) \le -1/2\}$$

Split S_p into $\tilde{b}$ disjoint parts $\tilde{\Pi}_1, \ldots, \tilde{\Pi}_{\tilde{b}}$, such that $\forall\ x, y \in \tilde{\Pi}_j$, $x'y \ge 1/2$. Let $\tilde{\xi}_j \in \tilde{\Pi}_j$, $j = 1, \ldots, \tilde{b}$. Define $\bar{F}^j_n$ as $\bar{F}^j_n = \{i \le n,\ \ell(x_i)'\tilde{\xi}_j \ge 275/280\}$, $j = 1, \ldots, \tilde{b}$. By Assumption 3, there exists $\tilde{\delta}_2 > 0$ such that

$$\#(\bar{F}^j_n) \ge \tilde{\delta}_2 n, \qquad j = 1, \ldots, \tilde{b}$$

It is easy to see that $-\ell(\bar{\beta}_{kn} - \beta_0) \in \tilde{\Pi}_j$ and $i \in \bar{F}^j_n$ imply that

$$\ell(x_i)'\ell(\bar{\beta}_{kn} - \beta_0) \le -1/2, \qquad \text{i.e. } i \in \bar{E}_n$$

For $i \in \bar{E}_n$, we have

$$|x_i'(\bar{\beta}_{kn}-\beta_0)| = |x_i||\bar{\beta}_{kn}-\beta_0||\ell(x_i)'\ell(\bar{\beta}_{kn}-\beta_0)| \geq \tilde{m}\tilde{\eta}/2$$

Hence

$$\bar{Q}_n(\bar{\beta}_{kn}) \geq \max_{i \varepsilon \bar{E}_n} e_i + \tilde{m}\tilde{\eta}/2$$

Thus

$$P(\bar{Q}_n(\bar{\beta}_{kn}) \leq a_2 + \tilde{m}\tilde{\eta}/4)$$

$$\leq P(\max_{i \varepsilon \bar{E}_n} e_2 \leq a_2 - \tilde{m}\tilde{\eta}/4)$$

$$\leq \sum_{j=1}^{\tilde{b}} P(\max_{i \varepsilon \bar{E}_n} e_i \leq a_2 - \tilde{m}\tilde{\eta}/4,\ -\ell(\bar{\beta}_{kn} - \beta_0)\ \varepsilon\ \tilde{\Pi}_j)$$

$$\leq \sum_{j=1}^{\tilde{b}} P(\max_{i \varepsilon \bar{F}_n^j} e_i \leq a_2 - \tilde{m}\tilde{\eta}/4, -\ell(\bar{\beta}_{kn}-\beta_0)\ \varepsilon\ \tilde{\Pi}_j)$$

$$\leq \sum_{j=1}^{\tilde{b}} P(\max_{i \varepsilon \bar{F}_n^j} e_i \leq a_2 - \tilde{m}\tilde{\eta}/4)$$

$$\leq \tilde{b}(1 - \tilde{\Delta}\tilde{m}\tilde{\eta}/4)^{\tilde{\delta}_2 n} \leq \tilde{b}/n^2$$

for large n. By Borel-Cantelli Lemma, we have with probability one, when n large enough

$$\bar{Q}_n(\bar{\beta}_{kn}) \geq a_2 + \tilde{m}\tilde{\eta}/4.$$

Thus for $k < k_0$, we have

$$\bar{R}_k - \bar{R}_{k_0} = \bar{Q}_n(\bar{\beta}_{kn}) - \bar{Q}_n(\bar{\beta}_{k_0 n}) - (k_0 - k)C_n$$

$$\leq \tilde{m}\tilde{\eta}/4 - (k_0 - k)C_n > 0, \tag{5}$$

for largh n, since $C_n \to 0$.

(4) and (5) proves $\bar{k}$ is consistent. Summarize the above arguments, we get the following theorem.

Theorem 2. Choose C_n satisfying

(i) $C_n \to 0$,

(ii) $nC_n/\log n \to \infty$.

Suppose the four assumptions given at the beginning of this section are true, then $\bar{k} \to k$, a. s.

Proof. Use the arguments given before, we only need to notice that for any sequence of C_n satisfying (i) and (ii), we can always choose r_n such that

(i)' $r_n/C_n \to 0$

(ii)' $nr_n/\log n \to \infty$

Q.E.D.

4. General Case

In this section we consider the same regression model (1). But the problem we are going to solve is to determine the subset (or the model) $J = \{1 \le j_1 < \dots < j_k \le p\}$ such that $\beta_j \ne 0$ if and only if $j \in J$. We make the same assumptions as given in previons sections.

Of course, we can use the procedure described in section 2 and 3 to determine the model J as follows: For each permutation π of $\beta = (\beta_1, \dots, \beta_p)'$, similarly rearranging $(x_{1i}, \dots, x_{pi})'$, we get a new model M_π. Under this model, using the approach given in section 2 and 3, we obtain estimates $\hat{k} = \hat{k}_{\hat{\pi}} = \min_\pi \hat{k}_\pi$ and $\bar{k} = \bar{k}_{\bar{\pi}} = \min_\pi \bar{k}_\pi$ and let $\hat{J}_1 = \{\hat{\pi}(1), \dots, \hat{\pi}(\hat{k})\}$ and $\bar{J}_1 = \bar{J}_1 = \{\bar{\pi}(1), \dots, \bar{\pi}(\bar{k})\}$, we can easily prove that, by using Theorem 1 and 2, $\hat{J}_1 \to J$, a. s. and $\bar{J}_1 \to J$, a. s.

An alternative method to estimate J is given as follows: Suppose T is a subset of $\{1, \dots, p\}$. Consider the model T:

$$y_n = x_n(T)'\beta(T) + e_n,$$

where $x_j(T) = (x_{ji}, j \in T)'$ and $\beta(T) = (\beta_j, j \in T)'$. Let

$$Q_n(T) = \min_{\beta(T)} \{ \max_{1 \le i \le n} (y_i - x_i(T)'\beta(T)) - \min_{1 \le i \le n} (y_i - x_i(T)'\beta(T)) \}$$

and

$$\bar{Q}_n(T) = \min_{\beta(T)} \max_{1 \le i \le n} |y_i - x_i(T)'\beta(T)|.$$

Define

$$\hat{R}_T = Q_n(T) + \#(T)c_n$$

and

$$\bar{R}_{\hat{J}_2} = \bar{Q}(T) + \#(T)c_n$$

Choose $\hat{J}_2$ such that

$$\hat{R}_{\hat{J}_2} = \min_T \hat{R}_T$$

and choose $\bar{J}_2$ such that

$$\bar{R}_{\bar{J}_2} = \min_T \bar{R}_T$$

We can also prove that $\hat{J}_2 \to J$, a. s. and $\bar{J}_2 \to J$, a. s. . However, there would be too much computation involved when p is relatively large. In the first case, there are totally p! permutations whileas in the second there are 2^p subsets of $\{1, \ldots, p\}$. In light of this, we propose another approach to estimate J which only involves p + 1 quantities to be computed.

Now let

$$\beta(j) = (\beta_1, \ldots, \beta_{j-1}, 0, \beta_{j+1}, \ldots, \beta_p)'$$

and define

$$Q_n(j) = \min_{\beta(j)} \{ \max_{1 \le i \le n} (y_i - x_i'\beta(j))$$

$$- \min_{1 \le i \le n} (y_i - x_i'\beta(j))\}$$

and

$$\bar{Q}_n(j) = \min_{\beta(j)} \max_{1 \le i \le n} |y_i - x_i' \beta(j)|.$$

Write

$$\hat{R}(n,j) = Q_n(j) - \hat{Q}_p - C_n$$

and

$$\bar{R}(n,j) = \bar{Q}_n(j) - \bar{Q}_p - C_n.$$

We choose

$$\hat{J}_n = \{\hat{j}_1, \ldots, \hat{j}_{\hat{k}_n}\} = \{j : \hat{R}(n,j) > 0\}$$

and

$$\bar{J}_n = \{\bar{j}_1, \ldots, \bar{j}_{\bar{k}_n}\} = \{j : \bar{R}(n,j) > 0\}$$

Then we have the following theorems.

Theorem 3. Under the conditions of theorem 1, we have that

$$\hat{J}_n \to J, \quad \text{a. s.}$$

where model $J = \{i_1, \ldots, i_k\}$ is the true one.

Proof If $j \in J$, by (3) with the replacement that $k_0 = p$ and $k = p-1$, we have that with probability one, $\hat{R}(n,j) > 0$ for all large n, i. e., $j \in \hat{J}_n$. Hence, when n

large enough, $\hat{J}_n$ J. Conversely, if $j \notin J$, using the same argument as proving theorem 1, we have

$$\hat{R}(n,j) = Q_n(j) - \hat{Q}_p - C_n$$

$$\leq 0(\log n/n) - C_n \quad \text{a. s.}$$

which together with (ii) implies that

$$\hat{R}(n,j) < 0, \qquad \text{for large } n,$$

i. e. $j \notin \hat{J}_n$ when n large enough. Therefore $\hat{J}_n$ J which completes the proof of Theorem 3.

Theorem 4. Under the conditions of theorem 2, we have that

$$\bar{J}_n \to J, \quad \text{a. s.}$$

where model $J = \{j_1, \ldots, j_k\}$ is the true one.

Proof. If $j \varepsilon J$, by (5) with the replacement that $k_0 = p$ and $k = p-1$, we have that with probability one, $\bar{R}(n,j) > 0$ for all large n, i. e., $j \varepsilon \bar{J}_n$. Hence, when n large enough, $\bar{J}_n$ J. Conversely, if $i \notin J$, using the same argument as proving theorem 2, we have

$$\bar{R}(n,j) = \bar{Q}_n(j) - \bar{Q}_p - C_n$$

$$\leq 0(\log n/n) - C_n \quad \text{a. s.}$$

which together with (ii) implies that

$$\bar{R}(n,j) < 0, \qquad \text{for large } n,$$

i. e. $j \notin \bar{J}_n$ when n large enough. Therefore $\bar{J}_n$ J which completes the proof of Theorem 4.

REFERENCES

1. Drygas, H. (1976). Weak and strong consistency of the lease squares estimators in regression models. Z. Wahrsch. Verw. Gebiete, **34**, 119-127

2. Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Stong consistency of least squares estimates in multiple regression II. Journal of Multivariate Analysis, **9**, 343-361

3. Oberhofer, W. (1982). The Consistency of nonlinear regression minimizing the L_1-norm. Ann. Statist. 10, 316-319

4. Wu, Y. (1986). On strong consistent estimates of regression coefficients when the errors are not independently and identically distributed. Tech. Report No. 86-06, Center for Multivariate Analysis, Univ. of Pittsburgh

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 87-15	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Strong consistency of estimation of number of regression variables when the errors are independent and their expectations are not equal to each other.		5. TYPE OF REPORT & PERIOD COVERED technical - June 1987 6. PERFORMING ORG. REPORT NUMBER 87-15
7. AUTHOR(s) Yuehua Wu		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis University of Pittsburgh, 515 Thackeray Hall Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332		12. REPORT DATE June 1987 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear model, model selection, regression coefficient, strong consistency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider the linear regression model $y_i = x_i'\beta + e_i$, $i = 1,2,\ldots$, where $\{x_i\}$ is a sequence of known p-vectors, $\beta' = (\beta_1,\ldots,\beta_p)$ is an unknown p-vector, known as regression coefficients, $\{e_i\}$ is a sequence of random errors. It is of interest to test the hypothesis H_k: $\beta_{k+1} = \ldots = \beta_p = 0$, $k = 0,1,\ldots,p$. We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support $[a_1,a_2]$. Under certain

DD FORM 1 JAN 73 1473

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

conditions, we obtain the strongly consistent estimate of the number of k for which $\beta_k \neq 0$ and $\beta_{k+1} = \ldots \beta_p = 0$, by using the information theoretical criteria.

END

DATE

FILMED

DEC.

1987